

Project: Capstone Project 1: Statistical Data Analysis

As per data, it is observed that tube assemblies are procured either in bulk or non-bulk (i.e. minimum order basis). I would like to learn the difference in buying patterns and see if there is any change in the assembly parameters such as weight, cost and quantity associated with the two pricing types.

Two-Sided T-Test:

To do this, I have used two-sided t-test to calculate t-statistics and p-values of multiple parameters. I noticed that p-value was < 0.05 , hence rejected null and proceeded with alternate hypothesis that both pricing types have different assembly parameters.

Please see result of hypothesis testing and t-test results for reference:

Test Hypothesis:

1. Null: Bulk and non-bulk assemblies have same parameters. There is no difference in weight, cost and quantity patterns.
2. Alternate: Non-bulk assemblies parameters differ from bulk assemblies. There is difference in weight, cost and quantity patterns.

Result of t-test:

	t-statistics	p_value
parameters		
weight	-14.843691	1.360478e-49
annual_usage	-15.864254	2.373634e-56
min_order_quantity	-50.218847	0.000000e+00
quantity	18.082729	1.539973e-72
%supplier_rebate	82.783817	0.000000e+00
extended_cost	11.419592	4.092552e-30
total_cost	-11.578067	6.601705e-31
cost	-8.984690	2.807771e-19

Bootstrap Sampling:

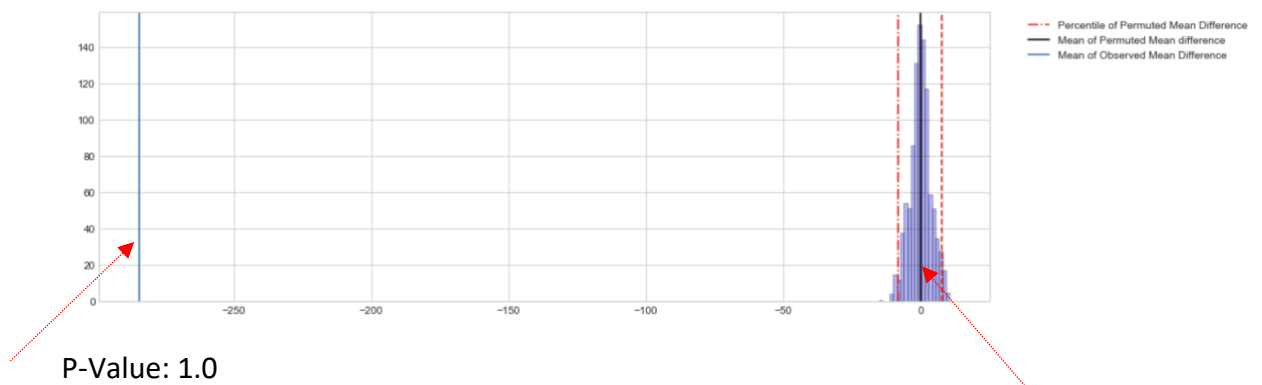
However, this is just one sample we have used to drive inference. We would be collecting data on a regular basis. Hence we want to find out, could there be an instance when bulk parameters change so significantly

that business has to do bulk buying on a minimum order basis. Would that be possible, and if yes how much shift in the p-value?

To investigate this, I used bootstrap sampling to simulate and calculated p-values between observed mean difference and permuted mean difference for each pricing types, using 1000 samples.

Test Hypothesis:

1. Null: No change in Business Trend. Bulk assembly parameters differ from non-bulk assemblies as per above. Business will buy bulk assemblies in bulk.
2. Alternate: Business Trend has changed. Bulk and non-bulk assemblies share some common parameters. Business will buy bulk assemblies as non-bulk (i.e. minimum order basis)



Confidence Interval of Mean of Permuted Mean Difference at 95% level:
[-8.10, 7.93]

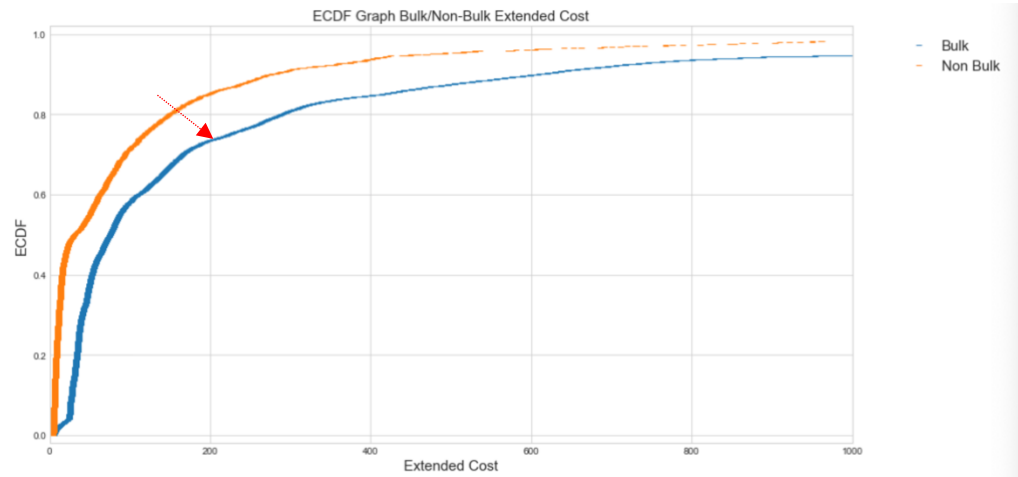
Mean of Observed Mean Difference: -284.16

Since, $p\text{-value} > 0.05$, null exists. Business will continue to buy bulk assemblies in bulk.

However, any shift in business trend from above would require increase in the mean of observed mean difference from -284 to upper sample range of 7.93 in order to buy bulk assemblies as non-bulk (i.e. minimum order basis). This would require significant change in bulk assembly parameters such as weight, costs, quantity which can be tracked and compared with the collected data on ongoing basis.

Empirical Cumulative Distribution Function (ECDF):

To further understand the significance of this shift in business trend from



bulk to minimum order quantity, we will use ECDF graph to measure variance between two extended costs.

This shows two key observations:

1. Since, both ECDF's do not overlay, bulk and non-bulk extended cost have variance.
2. Bulk extended cost is trending lower than non-bulk.

-----End-----